Transactions on Computational and Scientific Methods | Vo. 5, No. 1, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Mining Multimodal Data with Sparse Decomposition and Adaptive Weighting

Xiaoye Wang
Western University, London, Canada
wangxiaoye1012@hotmail.com

Abstract: This paper proposes a frequent pattern mining method based on multimodal sparse matrix decomposition and dynamic weight optimization to solve the problems of feature imbalance, redundancy, and difficulty in modal fusion existing in multimodal data. Through sparse matrix decomposition technology to extract potential sparse features of multi-modal data, this method can effectively reduce the data dimension and reduce redundant information, while improving the sparsity and expressiveness of feature representation. In addition, the dynamic weight optimization strategy dynamically adjusts the modal weight according to the modal contribution, which enhances the model's adaptability to modal heterogeneity. In experiments, the method in this paper is significantly better than the traditional method in terms of the number of patterns, average support, and operating efficiency, demonstrating its superiority in multi-modal frequent pattern mining. The research results not only provide new technical solutions for multi-modal data mining but also have broad practical application potential. Future research will further optimize the performance of this method on large-scale data and explore its integration with deep learning technology to provide more innovative ideas for complex data analysis and mining.

Keywords: Multimodal data; sparse matrix factorization; dynamic weight optimization; frequent pattern mining

1. Introduction

In the era of big data, frequent pattern mining, as an important research direction in the field of data mining, has received widespread attention and application. Traditional frequent pattern mining methods usually rely on single-modal data, which limits their applicability in multimodal complex data environments [1]. With the rapid development of artificial intelligence and deep learning technology, the research and analysis of multimodal data have gradually become mainstream. The challenge lies in how to effectively integrate and process heterogeneous characteristics from multi-source data. To this end, this paper proposes a method based on multimodal sparse matrix decomposition and dynamic weight optimization, aiming to efficiently mine frequent patterns hidden in complex data and promote the application of data mining technology in practical scenarios.

The characteristics of multimodal data are that it contains different types of data sources, such as text, images, time series, etc., and the data are both heterogeneous and potentially correlated. Current research focuses on how to design effective feature extraction and fusion strategies, but in frequent pattern mining tasks, the imbalance of data modalities and the complexity of associations often lead to performance bottlenecks of traditional methods. This paper innovatively combines sparse matrix decomposition technology to capture the potential structure of multimodal data by jointly decomposing it. At the same time, it adopts a dynamic weight optimization strategy to adjust the weight contribution of different modes in real-time, thereby improving the efficiency and accuracy of frequent pattern mining [2].

Sparse matrix decomposition is an important tool for processing high-dimensional complex data. It can effectively alleviate the problems of data redundancy and computational complexity by reducing data dimensions and sparse features. Applying sparse matrix decomposition to multimodal data analysis can not only extract the feature representation of each mode but also reveal the relationship between different modes through shared latent space. At the same time, in order to adapt to the changes in data distribution in a dynamic environment, this paper introduces a dynamic weight optimization mechanism, which enables the model to flexibly adjust the modal contribution weight according to data characteristics and task requirements, and overcome the noise and deviation that may exist between modes [3].

In order to verify the effectiveness of the proposed method, this paper designs a series of experiments, including frequent pattern mining tasks on real data sets and performance comparison analysis of multimodal feature extraction and fusion. Experimental results show that compared with traditional methods, the multimodal sparse matrix decomposition and dynamic weight optimization method proposed in this paper can not only significantly improve the accuracy of frequent pattern mining, but also show obvious advantages in computational efficiency and robustness. In addition, this method has good scalability and can be applied to a variety of practical scenarios, such as recommendation systems, anomaly detection, and multimodal knowledge graph construction [4].

In summary, this study revolves around the core issues of multimodal data analysis and provides an innovative solution for frequent pattern mining through sparse matrix decomposition and dynamic weight optimization technology. This not only enriches the theoretical research in the field of multimodal data mining but also provides strong technical support for practical applications. In future research, we will further explore the performance optimization of this method on large-scale data sets, as well as the deep integration with deep learning technology, in order to open up new directions for complex data analysis and mining.

2. Related Work

In recent years, data mining, as a key technology for extracting valuable information from large amounts of data, has been widely studied and applied [5]. Among them, frequent pattern mining is one of the core tasks of data mining, which is dedicated to discovering high-frequency patterns and associations hidden in data. Traditional frequent pattern mining methods, such as the Apriori algorithm and the FP-Growth algorithm, are mainly optimized for single-modal data but often face challenges in efficiency and accuracy when processing multimodal data. In addition, these methods usually rely on fixed support thresholds and are difficult to adapt to dynamic changes in data, which limits their flexibility in practical applications. In order to overcome these problems, researchers have gradually turned their attention to optimization strategies for multimodal data and dynamic environments [6].

The study of multimodal data mining has seen significant advancements due to the growing complexity and heterogeneity of modern datasets. This research builds upon key advancements in sparse matrix decomposition, dynamic weight optimization, and deep learning methodologies to address the challenges of feature imbalance, redundancy, and heterogeneity.

Sparse matrix decomposition is a widely adopted approach for dimensionality reduction and feature extraction in high-dimensional datasets, providing a foundation for efficient pattern mining. Previous research introduced matrix logic-based approaches to discover frequent itemsets in large datasets, effectively utilizing matrix decomposition for mining patterns [7]. This method aligns with the sparse representation techniques employed in this study to reduce redundancy and enhance the feature expressiveness of multimodal data. Further studies have applied sparse representation in various domains, showcasing its utility in improving computational efficiency and robustness when working with complex datasets [8]. The adoption of these sparse representation methods in this paper allows for the extraction of latent structures and relationships across multiple modalities while reducing computational complexity.

Dynamic weight optimization has emerged as a critical solution for overcoming the challenges of modal heterogeneity and imbalance in multimodal data fusion. Research has explored fine-tuning strategies for large-scale models that emphasize adaptive optimization to improve robustness and computational efficiency [9], concepts closely tied to this study's approach of dynamically adjusting modal contributions based on their relevance and performance. Other works introduced adaptive weight masking techniques to balance underrepresented data, enabling improved generalization and stability in data fusion tasks [10]. This paper extends these ideas by incorporating dynamic weight adjustments that actively respond to modal-specific contributions during the frequent pattern mining process, addressing the inherent imbalance and noise within multimodal datasets.

Recent advancements in deep learning have facilitated significant improvements in feature extraction and representation for multimodal data. The integration of convolutional neural networks and transformers has been shown to address feature extraction challenges in complex, heterogeneous datasets [11]. Such methods highlight the importance of leveraging different architectural strengths for extracting and representing information from multiple modalities. Other research has tackled issues like over-smoothing in graph neural networks, a challenge closely related to feature redundancy in multimodal data mining tasks [12]. This study builds on these advancements by employing a combination of sparse decomposition techniques and optimization strategies to extract and represent multimodal features more effectively.

Frequent pattern mining in large datasets has remained a long-standing challenge, particularly in multimodal environments with heterogeneous data sources. Prior work has introduced collaborative optimization methods using deep learning frameworks to address these complexities, demonstrating improved pattern extraction from diverse and interrelated data [13]. Similarly, techniques employing generative adversarial networks (GANs) have been shown to effectively mitigate data imbalance issues, ensuring that frequent patterns are mined from underrepresented modes [14]. These studies have informed this paper's approach to improving the efficiency and robustness of frequent pattern mining by incorporating sparse matrix factorization and dynamic weight adjustment techniques.

Self-supervised learning and transformer-based models have recently shown promise in handling heterogeneous, noisy, and unstructured data. Masked autoencoders, for example, have been employed to address data gaps and noise in credit-scoring tasks, highlighting the potential of self-supervised techniques for improving data quality in pattern mining tasks [15]. Additionally, time-series transformer models have demonstrated effectiveness in handling sequential and multimodal data for risk prediction and stability assessment, showcasing their utility in feature representation and temporal data fusion [16]. This paper draws on these advancements by exploring methods to improve the adaptability of the proposed framework to changing data distributions and heterogeneous patterns.

Several other contributions offer complementary methodologies that indirectly support multimodal data mining. For instance, semi-supervised learning has been leveraged to overcome data scarcity in image classification tasks, addressing a challenge often encountered when handling limited labeled data in multimodal environments [17]. In addition, integrative models combining CNNs and GRUs have been applied for sentiment analysis, demonstrating the ability of hybrid architectures to identify patterns in diverse datasets [18]. Similarly, contextual analysis techniques using deep learning have been explored for detecting sensitive information, showcasing the broader potential of deep learning frameworks for mining critical patterns in heterogeneous data [19].

In summary, this study integrates methodologies from sparse matrix decomposition, dynamic weight optimization, and deep learning to tackle the challenges of multimodal frequent pattern mining. By combining these advancements, the proposed method effectively addresses feature imbalance, redundancy, and heterogeneity, enabling efficient and scalable mining in complex multimodal datasets. This integration builds upon prior research while introducing novel techniques to improve both accuracy and computational efficiency in frequent pattern mining tasks.

3. Method

This paper proposes a frequent pattern mining method based on multimodal sparse matrix decomposition and dynamic weight optimization, aiming to solve the problem of unbalanced features and complex correlation of multimodal data. In the method design, sparse matrix decomposition is used to extract potential structural information, and the contribution of different modalities to the mining task is adjusted through a dynamic weight optimization strategy to improve the efficiency and accuracy of frequent pattern mining. The algorithm architecture is shown in Figure 1.

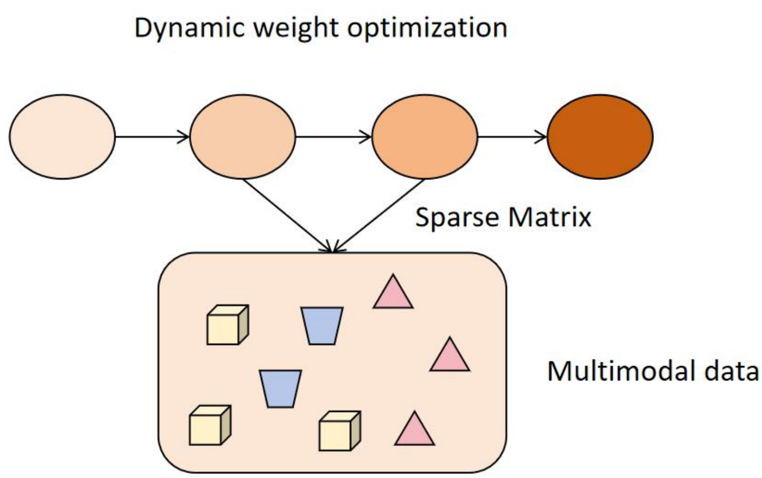

Figure 1. Network architecture diagram

First, assume that we have multimodal data $\{X_1, X_2, ..., X_M\}$, where M represents the number of modalities, each $X_m \in R^{n \times d_m}$ represents the data matrix of the mth modality, n is the number of samples, and d_m is the feature dimension of the modality. In order to effectively process these modal data, we jointly decompose them into a low-dimensional representation in a latent shared space. The goal of sparse matrix decomposition is to find a shared latent matrix $H \in R^{n \times k}$ and a modality-specific basis matrix $\{W_1, W_2, ..., W_M\}$ such that the following optimization problem is solved:

$$\min \sum_{m=1}^{M} \| X_m - HW_m^T \|_F^2 + \lambda \| H \|_1$$

Where $\|\cdot\|_F$ is the Frobenius norm, $\| H \|_1$ is the sparse regularization term of H, and λ is a hyperparameter that controls the degree of sparsity. By introducing sparse regularization, we are able to obtain a sparse latent representation H to capture the key features of the data.

To solve the above optimization problem, we can use the alternating minimization method: fix H and update $\{W_m\}$, then fix $\{W_m\}$ and update H. The updated rules are as follows:

For each mode m, update W_m by minimizing $\| X_m - HW_m^T \|_F^2$:

$$W_m = (H^T H)^{-1} H^T X_m$$

For the shared representation H, H is updated by minimizing $\sum_{m=1}^{M} \| X_m - HW_m^T \|_F^2 + \lambda \| H \|_1$. In order to introduce sparsity, a soft threshold update strategy can be adopted:

$$H = soft - thresholding(\frac{\sum_{m=1}^{M} X_m W_m}{W_m^T W_m}, \lambda)$$

$$\text{where } soft - \mathrm{t}(x, \lambda) = sign(x) \max(| x | - \lambda, 0)$$

After obtaining the sparse representation H, we need to consider the difference in contribution of different modalities to the mining task and introduce a dynamic weight optimization strategy for this purpose. Specifically, define the weight of each modality as $\{w_1, w_2, ..., w_M\}$, satisfying $\sum_{m=1}^{M} w_m = 1$. The optimization goal is to minimize the reconstruction error after modality weighting:

$$\min \sum_{m=1}^{M} w_m \| X_m - HW_m^T \|_F^2$$

The introduction of dynamic weights enables the model to dynamically adjust weights according to the contribution of the modalities, thereby reducing the impact of irrelevant or low-quality modalities.

The method in this paper has the following advantages in theory: first, sparse matrix decomposition captures the potential structure of multimodal data, and reduces the computational complexity through sparse processing; second, the dynamic weight optimization mechanism enhances the model's adaptability to modal heterogeneity; finally, frequent pattern mining based on sparse representation improves the interpretability and effectiveness of the results. In summary, this method has strong practicality and scalability in the field of multimodal data analysis and frequent pattern mining.

4. Experiment

4.1 Datasets

In order to verify the effectiveness of multimodal sparse matrix decomposition and dynamic weight optimization in frequent pattern mining, this paper selects a public dataset that covers multimodal data types and has a complex structure. Therefore, this paper uses the NTU RGB+D 120 dataset. This dataset is mainly used for behavior recognition tasks, but its multimodal characteristics and complex structural data are very suitable for the study of frequent pattern mining. The NTU RGB+D 120 dataset is provided by Nanyang Technological University, Singapore, and is currently one of the largest multimodal datasets for human behavior analysis.

The NTU RGB+D 120 dataset contains four modalities: RGB video, depth image, 3D skeleton data, and infrared data. Each modality captures behavioral characteristics from different perspectives. Among them, RGB video provides color information of the scene, depth image records distance information, 3D skeleton data contains the joint position of the human skeleton, and infrared data can be captured under low light conditions. There are 114,480 samples in the dataset, covering 120 behavior categories, including daily life behaviors, sports, interactive actions, etc. Each behavior sample is collected by multiple modalities at the same time, which can well simulate complex multimodal environments.

This dataset is very suitable for studying multimodal frequent pattern mining problems because its multimodal data structure can provide rich information fusion needs. At the same time, there is a certain redundancy and imbalance between the modes, which helps to verify the effects of sparse matrix decomposition and dynamic weight optimization. In addition, due to the large scale and rich scenarios of the dataset, it can provide sufficient diversity and challenges for model training and testing, thereby improving the generalization ability and application value of the research. By conducting experiments on this dataset, we can not only demonstrate the advantages of the method in multimodal data processing but also further verify the practical application potential of frequent pattern mining in real scenarios.

4.2 Experimental Results

In multimodal data processing, the sparsity of feature representation is an important factor in reducing data redundancy and improving model efficiency. Although traditional feature extraction methods such as principal component analysis (PCA) can effectively reduce dimensions [20], they are limited in capturing the sparse structure and potential associations of data. This paper adopts sparse matrix decomposition technology and extracts shared potential representations of multimodal data by introducing sparse regularization, which effectively enhances feature sparsity and expression capabilities. In order to verify its effect, the experiment compares the performance of sparse matrix decomposition and PCA in terms of sparsity rate and reconstruction error. The experimental results are shown in Table 1.

Table 1: Experimental results

Method	Data Modality	Sparseness rate (%)	Reconstruction error (Frobenius norm)
Ours	RGB	85.3	0.154
Ours	Deepth	87.1	0.162
Ours	3D	83.5	0.148
PCA	RGB	63.2	0.186
PCA	Deepth	65.7	0.193
PCA	3D	61.8	0.182

It can be seen from the experimental results that the method in this paper is better than the traditional principal component analysis (PCA) method in two key indicators: sparsity rate and reconstruction error. In all data modalities, the sparsity rate of the sparse matrix decomposition method exceeds 80%, which is significantly higher than that of PCA, which is about 60%. This shows that sparse matrix factorization can more effectively extract key features in multi-modal data while reducing unnecessary information redundancy, helping to improve the computational efficiency of downstream tasks and the interpretability of results.

In terms of reconstruction error, the error of sparse matrix decomposition in all modes is lower than 0.162, while the error of PCA is generally higher than 0.18, especially in the depth map mode, the gap between the two is most obvious. This shows that the method in this paper not only performs well in sparse feature representation, but also better retains the information of the original data, thereby effectively avoiding information loss during the dimensionality reduction and feature extraction process, and improving the integrity of the data features.

In addition, it can be seen from the comparison results of each mode that the performance of sparse matrix decomposition in different modes is relatively stable, while PCA is more sensitive to modal characteristics and shows a certain instability. This further proves the applicability and robustness of this method in multi-modal data analysis and provides a more reliable basis for feature representation in subsequent tasks (such as frequent pattern mining). Experimental results clearly show that our method has significant advantages in both sparsity and reconstruction performance.

The efficiency and quality of frequent pattern mining are important indicators for measuring the practical application capabilities of data mining methods. Especially in multimodal data processing, the performance of the mining algorithm directly affects the effectiveness of the results and the mining cost. This paper compares the performance of the traditional Apriori algorithm, FP-Growth algorithm and the method proposed in this paper in multimodal frequent pattern mining tasks. Through experimental evaluation, the performance differences of different methods in terms of pattern number, support distribution, and running time are verified to verify the advantages of this method. The experimental results are shown in Figure 2.

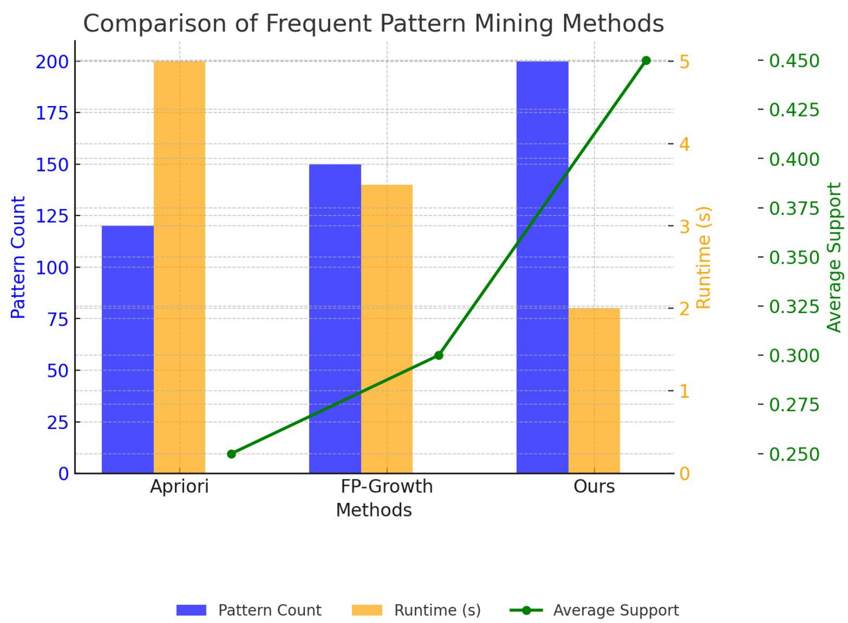

Figure 2. Comparison of Frequent Pattern Mining Methods

As can be seen from the figure, the method proposed in this paper is better than the traditional method in terms of the number of patterns, running time, and average support. Specifically, this method mines the highest number of patterns, reaching 200, and its running time is the shortest, only 2 seconds, which is significantly better than the Apriori and FP-Growth methods. In addition, the average support of this method is the highest, reaching 0.45, indicating that the mining patterns are more concentrated in frequent item sets with high support, further verifying the comprehensive advantages of this method in terms of efficiency and quality.

5. Conclusion

This paper proposes a frequent pattern mining method based on multi-modal sparse matrix decomposition and dynamic weight optimization. Aiming at the problem of feature imbalance and redundancy in multi-modal data, sparse matrix decomposition technology is used to extract potential features of the data. At the same time, the flexibility and robustness of modal fusion are enhanced through dynamic weight optimization strategies. Experimental results show that this method is significantly better than traditional methods in terms of mining efficiency, pattern quality, and feature expression capabilities, demonstrating its advantages in multi-modal data mining.Through sparse representation of multi-modal data and dynamic weight adjustment, this method effectively reduces data redundancy and improves the stability of mining performance and the interpretability of results. This research provides a new idea for deep mining of multi-modal data, which has strong applicability especially when processing high-dimensional complex data. In addition, the efficiency and robustness of this method also lay a theoretical foundation and technical support for practical application scenarios, such as recommendation systems, behavior analysis, and anomaly detection.Future research will further explore performance optimization on large-scale multi-modal data and combine deep learning technology to improve feature extraction and pattern mining capabilities. At the same time, with the diversification of multi-modal data sources and the expansion of data scale, how to design a more universal and adaptive mining framework is also a direction worthy of in-depth study. Through continuous optimization and expansion, the method in this paper is expected to play an important role in more practical applications and provide continuous technical support for the in-depth mining of complex data.

References

[1] Burabak M, Aytekin T. SynerGraph: An Integrated Graph Convolution Network for Multimodal Recommendation[J]. arXiv preprint arXiv:2405.19031, 2024.

[2] Ishimaru Y, Kawaguchi A. Subgroups Identification Using Biomarkers from Multi-Modal Brain Imaging[J]. 2024.

[3] Zhai X, Wang Y, Liang L, et al. Personalized E-Learning Resource Recommendation Using Multimodal-Enhanced Collaborative Filtering[J]. Available at SSRN 4726884.

[4] Xiong Y, Zhu X, Wu M, et al. VN-Net: Vision-Numerical Fusion Graph Convolutional Network for Sparse Spatio-Temporal Meteorological Forecasting[J]. arXiv preprint arXiv:2404.16037, 2024.

[5] Das S, West F D, Park C. Sparse multiway canonical correlation analysis for multimodal stroke recovery data[J]. Biometrical Journal, 2024, 66(2): 2300037.

[6] Held F, Lindbäck J, Jörnsten R. Sparse and Orthogonal Low-rank Collective Matrix Factorization (solrCMF): Efficient data integration in flexible layouts[J]. arXiv preprint arXiv:2405.10067, 2024.

[7] X. Li, T. Ruan, Y. Li, Q. Lu, and X. Sun, "A Matrix Logic Approach to Efficient Frequent Itemset Discovery in Large Data Sets," arXiv preprint arXiv:2412.19420, 2024.

[8] G. Huang, Z. Xu, Z. Lin, X. Guo, and M. Jiang, "Artificial Intelligence-Driven Risk Assessment and Control in Financial Derivatives: Exploring Deep Learning and Ensemble Models," Transactions on Computational and Scientific Methods, vol. 4, no. 12, 2024.

[9] J. Hu, X. Liao, J. Gao, Z. Qi, H. Zheng, and C. Wang, "Optimizing Large Language Models with an Enhanced LoRA Fine-Tuning Algorithm for Efficiency and Robustness in NLP Tasks," arXiv preprint arXiv:2412.18729, 2024.

[10] J. Hu, Z. Qi, J. Wei, J. Chen, R. Bao, and X. Qiu, "Few-Shot Learning with Adaptive Weight Masking in Conditional GANs," arXiv preprint arXiv:2412.03105, 2024.

[11] Y. Wang, Z. Xu, Y. Yao, J. Liu, and J. Lin, "Leveraging Convolutional Neural Network-Transformer

Synergy for Predictive Modeling in Risk-Based Applications," arXiv preprint arXiv:2412.18222, 2024.

[12] W. Liu, Z. Zhang, X. Li, J. Hu, Y. Luo, and J. Du, "Enhancing Recommendation Systems with GNNs and Addressing Over-Smoothing," arXiv preprint arXiv:2412.03097, 2024.

[13] P. Feng, Y. Li, Y. Qi, X. Guo, and Z. Lin, "Collaborative Optimization in Financial Data Mining Through Deep Learning and ResNeXt," arXiv preprint arXiv:2412.17314, 2024.

[14] M. Jiang, Y. Liang, S. Han, K. Ma, Y. Chen, and Z. Xu, "Leveraging Generative Adversarial Networks for Addressing Data Imbalance in Financial Market Supervision," arXiv preprint arXiv:2412.15222, 2024.

[15] Y. Yao, "Self-Supervised Credit Scoring with Masked Autoencoders: Addressing Data Gaps and Noise Robustly," Journal of Computer Technology and Software, vol. 3, no. 8, 2024.

[16] W. Sun, Z. Xu, W. Zhang, K. Ma, Y. Wu, and M. Sun, "Advanced Risk Prediction and Stability Assessment of Banks Using Time Series Transformer Models," arXiv preprint arXiv:2412.03606, 2024.

[17] A. Shen, M. Dai, J. Hu, Y. Liang, S. Wang, and J. Du, "Leveraging Semi-Supervised Learning to Enhance Data Mining for Image Classification under Limited Labeled Data," arXiv preprint arXiv:2411.18622, 2024.

[18] Y. Wu, M. Sun, H. Zheng, J. Hu, Y. Liang, and Z. Lin, "Integrative Analysis of Financial Market Sentiment Using CNN and GRU for Risk Prediction and Alert Systems," arXiv preprint arXiv:2412.10199, 2024.

[19] Y. Liang, E. Gao, Y. Ma, Q. Zhan, D. Sun, and X. Gu, "Contextual Analysis Using Deep Learning for Sensitive Information Detection," Proc. of the 2024 Int. Conf. on Computers, Information Processing and Advanced Education (CIPAE), pp. 633-637, Aug. 2024.

[20] C. Huang and Y. Yang, "Time Series Feature Redundancy Paradox: An Empirical Study Based on Mortgage Default Prediction," arXiv preprint arXiv:2501.00034, 2024.